A PROBLEM IN ESTIMATING A DISTRIBUTION BY MINIMUM RISK

by

Mete Özer

United States
Naval Postgraduate School

THESIS

A PROBLEM IN ESTIMATING A DISTRIBUTION
BY MINIMUM RISK

by

Mete Özer

September 1970

A Problem in Estimating a Distribution by Minimum Risk

by

METE ÖZER
Lieutenant (junior grade), Turkish Navy
Turkish Naval Academy, 1965

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1970

ABSTRACT

It is common to estimate a distribution by means of a step function. Such estimates can be made continuous by connecting the points of the steps with straight line segments. In this paper the best estimator of this class is found for data which is uniformly distributed using minimum risk, then this risk is compared with those of the sample distribution function and the Pyke estimator.

TABLE OF CONTENTS

I . INTRODUCTION

In reference [1] , Prof. Read studies the properties of some estimators for a continuous distribution. In particular , he computes the risk function (using the assumption of the uniform distribution) of the estimator that connects the points ($X_r, \frac{r}{n+1}$) , where X_r is the r^{th} smallest observation , with straight line segments . This estimator is shown to be better than the sample distribution function in an asymptotic sense . It is called the Pyke estimator .

It is natural to ask the question , " How well can a distribution be estimated by a function that connect the points (X_r, C_r) with straight line segments?" The purpose of this paper is to find the sequence $\{C_r\}$ that minimizes the integrated risk function (using the uniform distribution) and to compare that risk function with those of the sample distribution function and the Pyke estimator.

II. DETERMINATION OF THE OPTIMUM COEFFICIENTS

Let $\{C_1, C_2, \ldots, C_{n+1}\}$ be an increasing sequence with $C_0 = 0$ and $C_{n+1} = 1$. A continuous function estimator for $F(x)$ can be defined by

$$(2.1) \quad H(x) = C_r + \frac{x - X_r}{X_{r+1} - X_r}(C_{r+1} - C_r) \qquad \text{for} ; X_r \leq x \leq X_{r+1} \quad r = 0, \ldots, n$$

where $X_1 < X_2 \ldots < X_n$ are the order statistics of a random sample from an absolutely continuous distribution function $F(x)$. It is convenient to assume that the population sampled is bounded and contained in a finite interval $[a, b]$. Thus, we can define $X_0 = a$, $X_{n+1} = b$ and the risk function (R) by

$$(2.2) \qquad R = E \int_a^b \left[F(x) - H(x) \right]^2 dF(x)$$

For convenience the interval $[a, b]$ can be taken to $[0, 1]$. The estimator $H(x)$ is defined piecewise according to which of the random intervals (X_r, X_{r+1}) contains x. Assuming the data come from a uniform distribution The joint p.d.f. of any two order statistics, say X_r and X_{r+1}, is as easily espressed. Using $X_r = u$ and $X_{r+1} = v$ the joint p.d.f. is

$$(2.3) \qquad f_{X_r,X_{r+1}}(u,v) = \frac{n!}{(r-1)!(n-r-1)!} u^{r-1}(1-v)^{n-r-1} \quad \text{for } 0 < r < n$$

and $0 < u < v < 1$

$$(2.4) \qquad f_{X_0,X_1}(0,v) = n(1-v)^{n-1} \qquad \text{for } r = 0 \text{ and } u = 0$$

$$(2.5) \qquad f_{X_n,X_{n+1}}(u,1) = nu^{n-1} \qquad \text{for } r = n \text{ and } v = 1$$

Let define

$$(2.6) \qquad \Delta_r = C_{r+1} - C_r$$

So that

$$(2.7) \qquad C_r = \sum_{j=0}^{r-1} \Delta_j$$

and for simplicity $f_{X_r,X_{r+1}}(u,v) = f_r(u,v)$

Then rewriting (2.2) again

$$(2.8) \qquad R = \sum_{r=0}^{n} \iiint_{0 \leq u \leq x \leq v \leq 1} \left[x - C_r - \frac{x-u}{v-u}\Delta_r \right]^2 f_r(u,v)\,du\,dv\,dx$$

Now, the problem is to find the C_j values that give the minimum risk.

Using classical optimization technique with Lagrangian form the problem is to

$$\text{Minimize} \quad \Phi = R - \lambda \left(\sum_{j=0}^{n} \Delta_j - 1 \right)$$

$$\text{Subject to} \quad \sum_{j=0}^{n} \Delta_j = 1$$

using λ for the Lagrange multiplier . Since

$$(2.9) \qquad \Phi = \sum_{r=0}^{n} \iiint_{0 \le u \le x \le v \le 1} \left[x - \sum_{j=0}^{r-1} \Delta_j - \frac{x-u}{v-u} \Delta_r \right]^2 f_r(u,v)\,du\,dv\,dx$$

$$- \lambda \left(\sum_{j=0}^{n} \Delta_j - 1 \right)$$

The partial derivative of Φ with respect to Δ_k is

$$\frac{\partial \Phi}{\partial \Delta_k} = \iiint \left[x - \sum_{j=0}^{k-1} \Delta_j - \frac{x-u}{v-u} \Delta_k \right] \frac{x-u}{v-u} f_k(u,v)\,du\,dv\,dx$$

$$- \sum_{r=k+1}^{n} \iiint \left[x - \sum_{j=0}^{r-1} \Delta_j - \Delta_r \frac{x-u}{v-u} \right] f_r(u,v)\,du\,dv\,dx + \frac{\lambda}{2} = 0$$

for $k = 1 , \quad . \quad . \quad . \quad , n - 1$

The partial derivative of Φ with respect to Δ_0 and using (2.4) is

$$\frac{\partial \Phi}{\partial \Delta_0} = \iint \left(x - \frac{x}{v} \Delta_0 \right) \frac{x}{v}\, n(1-v)^{n-1}\,dv\,dx$$

$$+ \sum_{r=1}^{n} \iiint \left(x - \sum_{j=0}^{r-1} \Delta_j - \Delta_r \frac{x-u}{v-u} \right) f_r(u,v)\,du\,dv\,dx + \frac{\lambda}{2} = 0$$

The partial derivative of Φ with respect to Δ_n and using (2.5) is

$$\frac{\partial \Phi}{\partial \Delta_n} = \int\int \left[x - \sum_{j=0}^{n-1} \Delta_j - \Delta_n \frac{x-u}{1-u} \right] \frac{x-u}{1-u} u^{n-1} dudx + \frac{\lambda}{2n} = 0$$

Let

$$g(u,v) = \frac{x-u}{v-u} \quad , \quad I_r(1) = \int\int_{\substack{0 \le u \le x \\ x \le v \le 1}} f_r(u,v)dudv$$

$$I_r(g) = \int\int_{\substack{0 \le u \le x \\ x \le v \le 1}} g(u,v)f_r(u,v)dudv \quad , \quad I_r(g^2) = \int\int_{\substack{0 \le u \le x \\ x \le v \le 1}} g^2(u,v)f_r(u,v)dudv$$

Using these notations , then the system of derivatives is

$$(2.10) \quad \frac{\partial \Phi}{\partial \Delta_k} = \int_0^1 xI_k(g)dx - \sum_{j=0}^{k-1} \Delta_j \int_0^1 I_k(g)dx - \Delta_k \int_0^1 I_k(g^2)dx + \sum_{r=k+1}^{n} \int_0^1 xI_r(1)dx$$

$$- \sum_{r=k+1}^{n} \sum_{j=0}^{r-1} \Delta_j \int_0^1 I_r(1)dx - \sum_{r=k+1}^{n} \Delta_r \int_0^1 I_r(g)dx + \frac{\lambda}{2} = 0$$

for $k = 1 , \ldots , n-1$

All terms in the above eguations reduce to Beta functions . The relationship between Beta and Gamma function

$$B(m,n) = \int_0^1 v^{m-1}(1-v)^{n-1}dv = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = \frac{(m-1)!(n-1)!}{(m+n-1)!}$$

is used heavily. We proceed to evaluate these terms .

Evaluation of

$$\int_0^1 xI_k(g)dx = \int\int\int x\frac{x-u}{v-u}u^{k-1}(1-v)^{n-k-1}\frac{n!}{(k-1)!(n-k-1)!}dudvdx$$

$$= \frac{n!}{(k-1)!(n-k-1)!}\left\{\frac{1}{3}\int\int\frac{v^3-u^3}{v-u}u^{k-1}(1-v)^{n-k-1}dudv\right.$$

$$\left.-\frac{1}{2}\int\int\frac{v^2-u^2}{v-u}u^{k}(1-v)^{n-k-1}dudv\right\}$$

$$= \frac{n!}{(k-1)!(n-k-1)!}\left\{\frac{1}{3}\left[\frac{1}{k}\int_0^1 v^{k+2}(1-v)^{n-k-1}dv\right.\right.$$

$$\left.+\frac{1}{k+1}\int_0^1 v^{k+2}(1-v)^{n-k-1}dv+\frac{1}{k+2}\int_0^1 v^{k+2}(1-v)^{n-k-1}dv\right]$$

$$\left.-\frac{1}{2}\left[\frac{1}{k+1}\int_0^1 v^{k+2}(1-v)^{n-k-1}dv+\frac{1}{k+2}\int_0^1 v^{k+2}(1-v)^{n-k-1}dv\right]\right\}$$

(2.11) $$= \frac{3k+4}{6(n+1)(n+2)}$$

Evaluation of

$$\int_0^1 xI_r(1)dx = \int\int\int xu^{r-1}(1-v)^{n-r-1} \frac{n!}{(r-1)!(n-r-1)!} dudvdx$$

$$= \frac{n!}{(r-1)!(n-r-1)!} \frac{1}{2} \left\{ \int\int (v^2-u^2)u^{r-1}(1-v)^{n-r-1}dudv \right\}$$

$$= \frac{n!}{(r-1)!(n-r-1)!} \frac{1}{2} \left\{ \frac{1}{r} \int_0^1 v^{r+2}(1-v)^{n-r-1}dv \right.$$

$$\left. - \frac{1}{r+2} \int_0^1 v^{r+2}(1-v)^{n-r-1}dv \right\}$$

$$(2.12) \qquad = \frac{r+1}{(n+1)(n+2)}$$

Evaluation of

$$\int_0^1 I_k(g)dx = \int\int\int \frac{x-u}{v-u} u^{k-1}(1-v)^{n-k-1} \frac{n!}{(k-1)!(n-k-1)!} dudvdx$$

$$= \frac{n!}{(k-1)!(n-k-1)!} \left\{ \frac{1}{2} \int\int \frac{v^2-u^2}{v-u} u^{k-1}(1-v)^{n-k-1}dudv \right.$$

$$\left. - \int\int u^k(1-v)^{n-k-1}dudv \right\}$$

$$= \frac{n!}{(k-1)!(n-k-1)!} \left\{ \frac{1}{2} \left[\frac{1}{k} \int_0^1 v^{k+1}(1-v)^{n-k-1} dv \right. \right.$$

$$\left. \left. + \frac{1}{k+1} \int_0^1 v^{k+1}(1-v)^{n-k-1} dv \right] - \frac{1}{k+1} \int_0^1 v^{k+1}(1-v)^{n-k-1} dv \right\}$$

$$(2.13) \qquad = \frac{1}{2(n+1)}$$

Evaluation of

$$\int_0^1 I_k(g^2)dx = \iiint \left(\frac{x-u}{v-u}\right)^2 u^{k-1}(1-v)^{n-k-1} \frac{n!}{(k-1)!(n-k-1)!} \, dudvdx$$

$$= \frac{n!}{(k-1)!(n-k-1)!} \left\{ \frac{1}{3} \iint \frac{v^3 - u^3}{(v-u)^2} u^{k-1}(1-v)^{n-k-1} dudv \right.$$

$$- \iint \frac{v^2 - u^2}{(v-u)^2} u^k (1-v)^{n-k-1} dudv$$

$$\left. + \iint \frac{v-u}{(v-u)^2} u^{k+1}(1-v)^{n-k-1} dudv \right\}$$

$$= \frac{n!}{(k-1)!(n-k-1)!} \quad \frac{1}{3} \left\{ \frac{1}{k} \int_0^1 v^{k+1}(1-v)^{n-k-1} dv \right.$$

$$\left. - \frac{1}{k+1} \int_0^1 v^{k+1}(1-v)^{n-k-1} dv \right\}$$

$$(2.14) \qquad = \frac{1}{3(n+1)}$$

Evaluation of

$$\int_0^1 I_r(1)dx = \iiint u^{r-1}(1-v)^{n-r-1} \frac{n!}{(n-r)!(n-r-1)!} dudvdx$$

$$= \frac{n!}{(n-r)!(n-r-1)!} \left\{ \frac{1}{r} \int_0^1 v^{r+1}(1-v)^{n-r-1} dv \right.$$

$$\left. - \frac{1}{r+1} \int_0^1 v^{r+1}(1-v)^{n-r-1} dv \right\}$$

(2.15) $$= \frac{1}{n+1}$$

Evaluation of

$\int_0^1 I_r(g)dx$ is same as evaluation of $\int_0^1 I_k(g)dx$

(2.16) $$\int_0^1 I_r(g)dx = \frac{1}{2(n+1)}$$

The use of these evaluations (2.11), (2.12), (2.13), (2.14), (2.15), (2.16) in (2.10), then the derivative with respect to Δ_k is

(2.17) $$\frac{\partial \phi}{\partial \Delta_k} = \frac{3k+4}{6(n+2)} - \frac{1}{2} \sum_{j=0}^{k-1} \Delta_j - \frac{1}{3} \Delta_k$$

$$+ \frac{1}{(n+2)} \sum_{r=k+1}^{n} (r+1) - \sum_{r=k+1}^{n} \sum_{j=0}^{r-1} \Delta_j - \frac{1}{2} \sum_{r=k+1}^{n} \Delta_r + \frac{\lambda(n+1)}{2}$$

$$= \frac{3k+4}{n+2} - 3C_k - 2(C_{k+1} - C_k) + 3(n+1) - \frac{3(k+2)(k+1)}{(n+2)} - 6 \sum_{r=k+1}^{n} C_r$$

$$- 3(1 - C_{k+1}) + 3\lambda(n+1) = 0$$

or after evaluation the derivative (2.10) is

(2.18) $$C_k - C_{k+1} + 6 \sum_{r=k+1}^{n} C_r = 3\lambda(n+1) + 3n - \frac{3k^2+6k+2}{n+2}$$

for; $k = 1, \ldots, n-1$

The derivative with respect to Δ_0 is

(2.19) $$\frac{\partial \Phi}{\partial \Delta_0} = \int\int (x - \frac{x}{v}\Delta_0)\frac{x}{v} n(1-v)^{n-1} \, dvdx$$

$$+ \sum_{r=1}^{n} \int\int\int (x - \sum_{j=0}^{r-1}\Delta_j - \Delta_r \frac{x-u}{v-u}) f_r(u,v)\,dudvdx + \frac{\lambda}{2} = 0$$

$$= n\int\int \frac{x^2}{v}(1-v)^{n-1} \, dvdx - n\Delta_0 \int\int (\frac{x}{v})^2 (1-v)^{n-1} dvdx$$

$$+ \sum_{r=1}^{n} \int x I_r(1) \, dx - \sum_{r=1}^{n} C_r \int I_r(1)dx - \sum_{r=1}^{n} \Delta_r \int I_r(g)dx + \frac{\lambda}{2} = 0$$

This can be reduced to

(2.20) $$- C_1 + 6 \sum_{k=1}^{n} C_k = 3\lambda(n+1) + 3n - \frac{2}{n+2}$$ for ; $k = 0$

The derivative with respect to Δ_n is

(2.21) $$\frac{\partial \Phi}{\partial \Delta_n} = \int\int \left[x - \sum_{j=0}^{n-1}\Delta_j - \Delta_n \frac{x-u}{1-u}\right] \frac{x-u}{1-u} u^{n-1} \, dudx + \frac{\lambda}{2n} = 0$$

can be reduced to

$$\frac{\partial \Phi}{d\Delta_n} = \frac{2}{n+2} - \Delta_n - 3\lambda(n+1) = 0$$

or

$$(2.22) \qquad c_n = 3\lambda(n+1) + \frac{n}{n+2} \qquad \text{for; } k=n \, .$$

so in general

$$(2.23) \qquad c_k - c_{k+1} + 6 \sum_{r=k+1}^{n} c_r = 3\lambda(n+1) + 3n - \frac{3k^2+6k+2}{n+2}$$

$$\text{for ; } \quad k = 0, \ldots, n$$

Using (2.23) , then system of eguations in matrix form is

$$\begin{bmatrix} 5 & 6 & 6 & 6 & \cdot & \cdot & \cdot & 6 & 0 \\ 1 & 5 & 6 & 6 & \cdot & \cdot & \cdot & 6 & 0 \\ 0 & 1 & 5 & 6 & \cdot & \cdot & \cdot & 6 & 0 \\ & & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot & & \cdot & \cdot \\ & & & & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 1 & 5 & 6 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 & 5 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \cdot \\ \cdot \\ \cdot \\ c_{n-1} \\ c_n \\ c_{n+1} \end{bmatrix} \begin{bmatrix} \theta - 2/(n+2) \\ \theta - 11/(n+2) \\ \theta - 26/(n+2) \\ \cdot \\ \cdot \\ \cdot \\ \theta - (3n^2-6n+2)/(n+2) \\ \theta - (3n^2-1)/(n+2) \\ \theta - (3n^2+6n+2)/(n+2) \end{bmatrix}$$

where $\theta = 3\lambda(n+1) + 3n$

After successive row subractions the matrix is

$$\begin{bmatrix} 4 & 1 & 0 & 0 & . & . & . & 0 & 0 \\ 1 & 4 & 1 & 0 & . & . & . & 0 & 0 \\ 0 & 1 & 4 & 1 & . & . & . & 0 & 0 \\ & & . & . & . & & & . & . \\ & & & . & . & & & . & . \\ & & & & . & & & . & . \\ 0 & 0 & . & . & . & 1 & 4 & 1 & 0 \\ 0 & 0 & . & . & . & 0 & 1 & 4 & 1 \\ 0 & 0 & . & . & . & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ . \\ . \\ . \\ C_{n-1} \\ C_n \\ C_{n+1} \end{bmatrix} = \begin{bmatrix} 9/(n+2) \\ 15/(n+2) \\ 21/(n+2) \\ . \\ . \\ . \\ (6n-3)/(n+2) \\ (6n+3)/(n+2) \\ \theta-(3n^2+6n+2)/(n+2) \end{bmatrix}$$

To solve this matrix equation is considered the linear second-order difference equation such that

(2.24) $$C_k + 4C_{k+1} + C_{k+2} = \frac{6k+9}{n+2}$$

with initial conditions $C_0 = 0$ and $C_{n+1} = 1$

The corresponding homogeneous equation of (2.24) is

(2.25) $$C_k + 4C_{k+1} + C_{k+2} = 0$$

has the auxiliary equation

$$m^2 + 4m + 1 = 0$$

The roots of this equation can be calculated using the quadratic formula. These roots are $A = -2 + \sqrt{3}$ and $B = -2 - \sqrt{3}$

so the general sotion of (2.25) is given by

(2.26) $$C_k = K_1A^k + K_2B^k + C_k^*$$

Trial solution of C_k^* is

$$C_k^* = D_0 + D_1 k$$

Using (2.24)

$$D_0 = \frac{1}{2(n+2)}, \qquad D_1 = \frac{1}{n+2} \qquad \text{then}$$

$$C_k^* = \frac{1 + 2k}{2(n+2)}$$

So the general solution is

(2.27) $$C_k = K_1 A^k + K_2 B^k + \frac{1+2k}{2(n+2)}$$

using initial conditions to determine the values of K_1 and K_2

$C_0 = 0$ in (2.27) implies

(2.28) $$K_1 + K_2 = -\frac{1}{2(n+2)}$$

$C_{n+1} = 1$ in (2.27) implies

(2.29) $$K_1 A^{n+1} + K_2 B^{n+1} = \frac{1}{2(n+2)}$$

Simultaneous solution of (2.28) and (2.29) are

$$K_1 = \frac{1 + B^{n+1}}{2(n+2)\,(A^{n+1} - B^{n+1})} \quad \text{and} \quad K_2 = -\frac{1 + A^{n+1}}{2(n+2)(A^{n+1} - B^{n+1})}$$

so

(2.30) $$C_k = \frac{1}{(n+2)} \left\{ k + \frac{1}{2} - \frac{(A^{n+1-k} - B^{n+1-k}) - (A^k - B^k)}{2(A^{n+1} - B^{n+1})} \right\}$$

for; $k = 0, \ldots, n+1$

using $\Delta_k = C_{k+1} - C_k$ then

$$\Delta_k = \frac{1}{(n+2)} \left\{ 1 - \frac{(1 - A)(A^{n-k} + A^k) - (1 - B)(B^{n-k} + B^k)}{2(A^{n+1} - B^{n+1})} \right\} \tag{2.31}$$

for; $k = 0, \ldots, n$

III. CALCULATION OF RISK

From equation (2.8) the risk equation is

$$R = \sum_{r=0}^{n} \iiint \left[(x - C_r) - (g\Delta_r)\right]^2 f_r(u,v)\,du\,dv\,dx \tag{3.1}$$

$$= \sum_{r=0}^{n} \left\{ \int_0^1 x^2 I_r(1)dx - 2C_r \int_0^1 xI_r(1)\,dx + C_r^2 \int_0^1 I_r(1)dx \right.$$

$$\left. - 2\Delta_r \int_0^1 xI_r(g)dx + 2C_r\Delta_r \int_0^1 I_r(g)dx + \Delta_r^2 \int_0^1 I_r(g^2)dx \right\}$$

Using,

$$\int_0^1 x^2 I_r(1)dx = \frac{(r+2)(r+1)}{(n+3)(n+2)(n+1)} \qquad r = 0, \ldots, n$$

$$\int_0^1 xI_r(1)\,dx = \frac{r+1}{(n+1)(n+2)} \qquad r = 0, \ldots, n$$

$$\int_0^1 I_r(1)dx = \frac{1}{n+1} \qquad r = 0, \ldots, n$$

$$\int_0^1 xI_r(g)dx = \frac{3r+4}{6(n+1)(n+2)} \qquad r = 0, \ldots, n$$

$$\int_0^1 I_r(g)dx = \frac{1}{2(n+1)} \qquad r = 0, \ldots, n$$

$$\int_0^1 I_r(g^2)dx = \frac{1}{3(n+1)} \qquad r = 0, \ldots, n$$

So

$$(n+1) R = \frac{n+1}{3} - \frac{3n+4}{3(n+2)} - \frac{2}{(n+2)} \sum_{r=1}^{n} rC_r - \frac{1}{n+2} \sum_{r=1}^{n} C_r$$

$$+ \frac{1}{3} \sum_{r=0}^{n} \Delta_r^2 + \sum_{r=0}^{n} C_{r+1} C_r$$

Calculation of $\sum_{r=0}^{n} \Delta_r^2$

Using (2.31)

$$\sum_{r=0}^{n} \Delta_r^2 = \frac{1}{(n+2)^2} \sum_{r=0}^{n} \left\{ 1 - \frac{1}{2(A^{n+1} - B^{n+1})} \left[(1-A)(A^{n-r} + A^r) - (1-B)(B^{n-r} + B^r \right] \right\}^2$$

and using

$$\sum_{r=0}^{n} (A^{n-r} + A^r) = 2 \sum_{r=0}^{n} A^r = 2 \frac{1 - A^{n+1}}{1 - A}$$

$$\sum_{r=0}^{n} (B^{n-r} + B^r) = 2 \sum_{r=0}^{n} B^r = 2 \frac{1 - B^{n+1}}{1 - B}$$

Similiarly

$$\sum_{r=0}^{n} (A^{2(n-r)} - A^{2r}) = 2 \sum_{r=0}^{n} A^{2r} = 2 \frac{1 - A^{2(n+1)}}{1 - A^2}$$

$$\sum_{r=0}^{n} (B^{2(n-r)} + B^{2r}) = 2 \sum_{r=0}^{n} B^{2r} = 2 \frac{1 - B^{2(n+1)}}{1 - B^2}$$

and

$$\frac{1}{1 - A^{-2}} = \frac{1}{1 - B^2} = \frac{1 - A}{6(1+B)}$$

$$\frac{1}{1 - B^{-2}} = \frac{1}{1 - A^2} = \frac{1 - B}{6(1 + A)}$$

$$\frac{1 - A}{1 + A} = - \frac{1 - B}{1 + B}$$

$$\frac{B^n}{1 + A} (1 - A^{2(n+1)}) = \frac{1}{1 + B} (B^{n+1} - A^{n+1})$$

then

$$\sum_{r=0}^{n} \Delta_r^2 = \frac{n+3}{(n+2)^2} + \frac{n+1}{(n+2)^2} \frac{(1-A)^2 A^n + (1 - B)^2 B^n - 12}{2(A^{n+1} - B^{n+1})^2} \tag{3.2}$$

$$-\sqrt{3} \frac{(A^{n+1} + B^{n+1}) + 2}{2(n+2)^2 (A^{n+1} - B^{n+1})}$$

Calculation of $\sum_{r=0}^{n} C_{r+1} C_r$

Using (2.30)

$$\sum_{r=0}^{n} C_{r+1} C_r = \frac{1}{(n+2)^2} \sum_{r=0}^{n} \left\{ (r^2 + 2r + \tfrac{3}{4}) + (r + \tfrac{1}{2}) \left[(n+2)C_{r+1} - r - \tfrac{3}{2} \right] \right.$$

$$+ \quad (r+\tfrac{3}{2}) \left[(n+2)\, C_r - r - \tfrac{1}{2} \right]$$

$$+ \frac{1}{4} \; \frac{\left[A^{n+1-r} - A^{r} - B^{n+1-r} + B^{r}\right]\left[A^{n-r} - A^{r+1} - B^{n-r} + B^{r+1}\right]}{(A^{n+1} - B^{n+1})^2}$$

here

$$\sum_{r=0}^{n} (r^2 + 2r + \tfrac{3}{4}) = \frac{n(n+1)(4n+14) + 9(n+1)}{12}$$

and

$$\sum_{r=0}^{n} \frac{1}{4} \; \frac{\left[A^{n+1-r} - A^{r} - B^{n+1-r} + B^{r}\right]\left[A^{n-r} - A^{r+1} - B^{n-r} + B^{r+1}\right]}{(A^{n+1} - B^{n+1})^2}$$

$$= \frac{1}{24(A^{n+1}-B^{n+1})} \Big[A^{2n+3} + B^{2n+3} - (6n+4)(A^{n+2} + B^{n+2}) - (A^{2n+1} + B^{2n+1})$$
$$-(6n+7)(A^{n} + B^{n}) + 8\,(n+1)\Big]$$

Then

$$(3.3) \quad \sum_{r=0}^{n} C_{r+1} C_r = -\frac{n(n+1)(4n+14) + 9(n+1)}{12(n+2)^2} + \frac{1}{n+2}\left[n + \tfrac{1}{2} + 2\sum_{r=1}^{n} rC_r + \sum_{r=1}^{n} C_r\right]$$

$$+ \frac{\left[A^{2n+3}+B^{2n+3} - (6n+4)(A^{n+2}+B^{n+2}) - (A^{2n+1}+B^{2n+1}) - (6n+7)(A^{n}+B^{n}) + 8(n+\right.}{24(n+2)^2(A^{n+1} - B^{n+1})^2}$$

Using (3.2) and (3.3) then

$$(3.4) \quad (n+1)\, R = \frac{2n^2+3n-1}{12(n+2)^2} - \frac{\sqrt{3}\,(A^{n+1} + B^{n+1} + 4)}{12(n+2)^2\,(A^{n+1}-B^{n+1})}$$

$$- \frac{1}{24(n+2)^2(A^{n+1}-B^{n+1})^2}\Big[(2n+3)(A^{n}+B^{n}) + 40\,(n+1)$$

$$+2n(A^{n+2} + B^{n+2}) + 8(n+1)\,(A^{n+1} + B^{n+1})\Big]$$

IV. COMPARISONS AND SUMMARY

From (3.17) page 14 of Prof. Read's paper

$$(n+1)^2 R(\hat{C}) = \int_0^1 (n-1)x(1-x)dx + 2(n+1)\int_0^1 \left\{ (1-x)h(x) + xh(1-x) \right\} dx$$

where

$$\hat{C} = \frac{r}{n+1} \qquad \text{and} \qquad h(x) = \int_0^x \left(\frac{x-u}{1-u} \right)^n udu$$

$$(n+1)^2 R(\hat{C}) = \frac{n-1}{6} + 4(n+1)\int_0^1 (1-x)h(x)dx$$

and evaluation of

$$\int_0^1 (1-x)h(x)dx = \int_0^1 (1-x)\int_0^x \left(\frac{x-u}{1-u}\right)^n ududx$$

$$= \iint_{0<u<x<1} (1-x)\left(\frac{x-u}{1-u}\right)^n ududx$$

let

$$y = \frac{x-u}{1-u} \qquad 0 \le y \le 1$$

then

$$x = u + (1-u)y \ , \quad 1-x = (1-u)(1-y) \quad \text{and} \quad dy = \frac{dx}{1-u}$$

so

$$(4.1) \qquad (n+1)R(\hat{C}) = \frac{n-1}{6(n+1)} + \frac{1}{3(n+2)(n+1)}$$

where

$R(\hat{C})$ is Pyke risk

Risk of sample cumulative distribution function R(SDF) is

$$R(SDF) = \frac{1}{n} \int_0^1 x(1-x)dx = \frac{1}{6n}$$

then

$$(n+1)R(SDF) = \frac{n+1}{6n} \tag{4.2}$$

Optimum risk R(OPT) is derived in this paper is

$$(n+1)R(OPT) = \frac{2n^2+3n-1}{12(n+2)^2} - \frac{\sqrt{3}(A^{n+1}+B^{n+1}+4)}{12(n+2)^2(A^{n+1}-B^{n+1})} - \frac{(2n+3)(A^n+B^n)+8(n+1)\ (A^{n+1}+B^{n+1})+2n(A^{n+2}+B^{n+2})+40(n+1)}{24(n+2)^2(A^{n+1}-B^{n+1})^2} \tag{4.3}$$

In this eguation last two terms may be omitted.Since they are close to zero for $n \geqq 6$

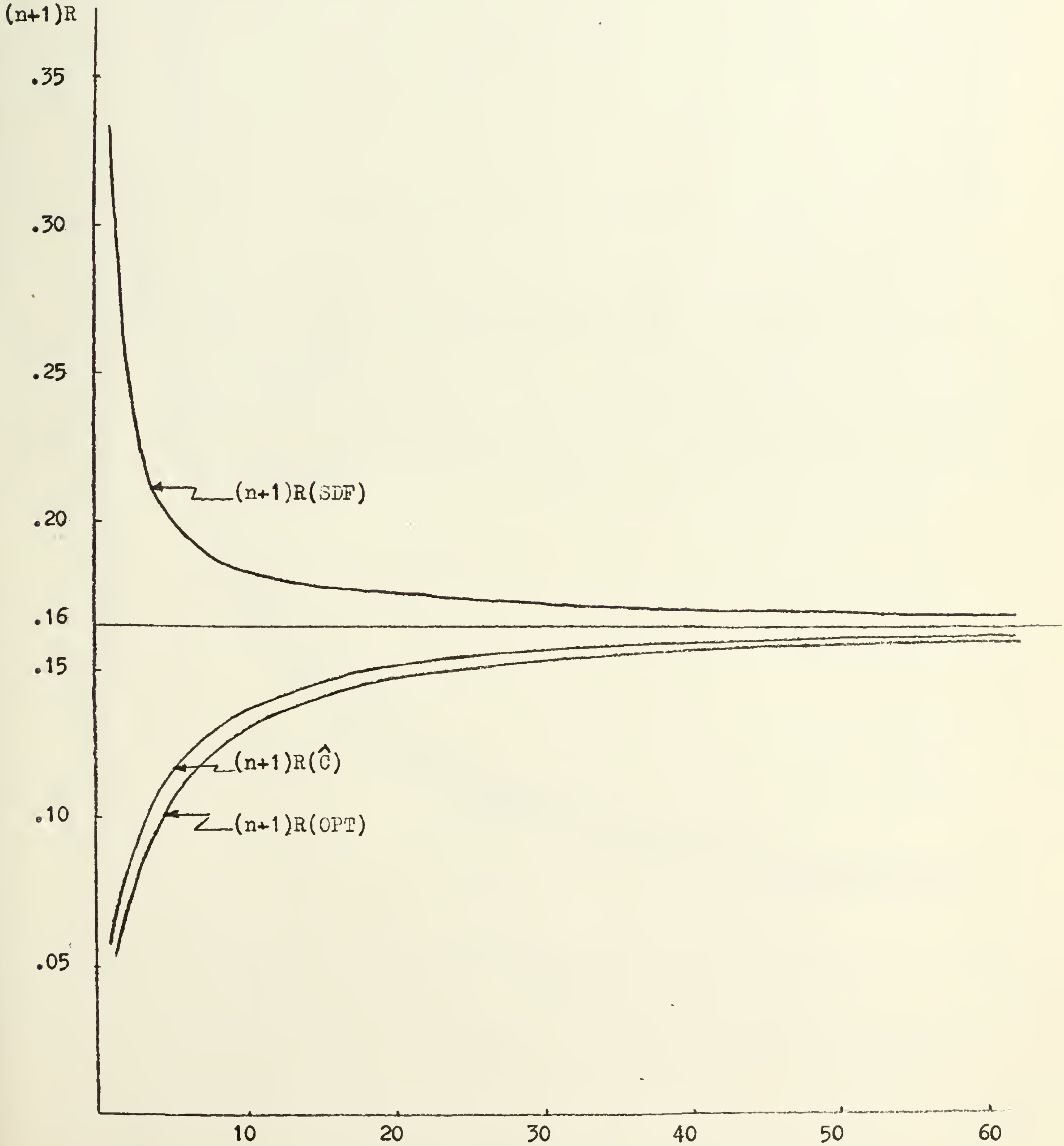
so

$$(n+1)R(OPT) \cong \frac{2n^2+3n-1}{12(n+2)^2}$$

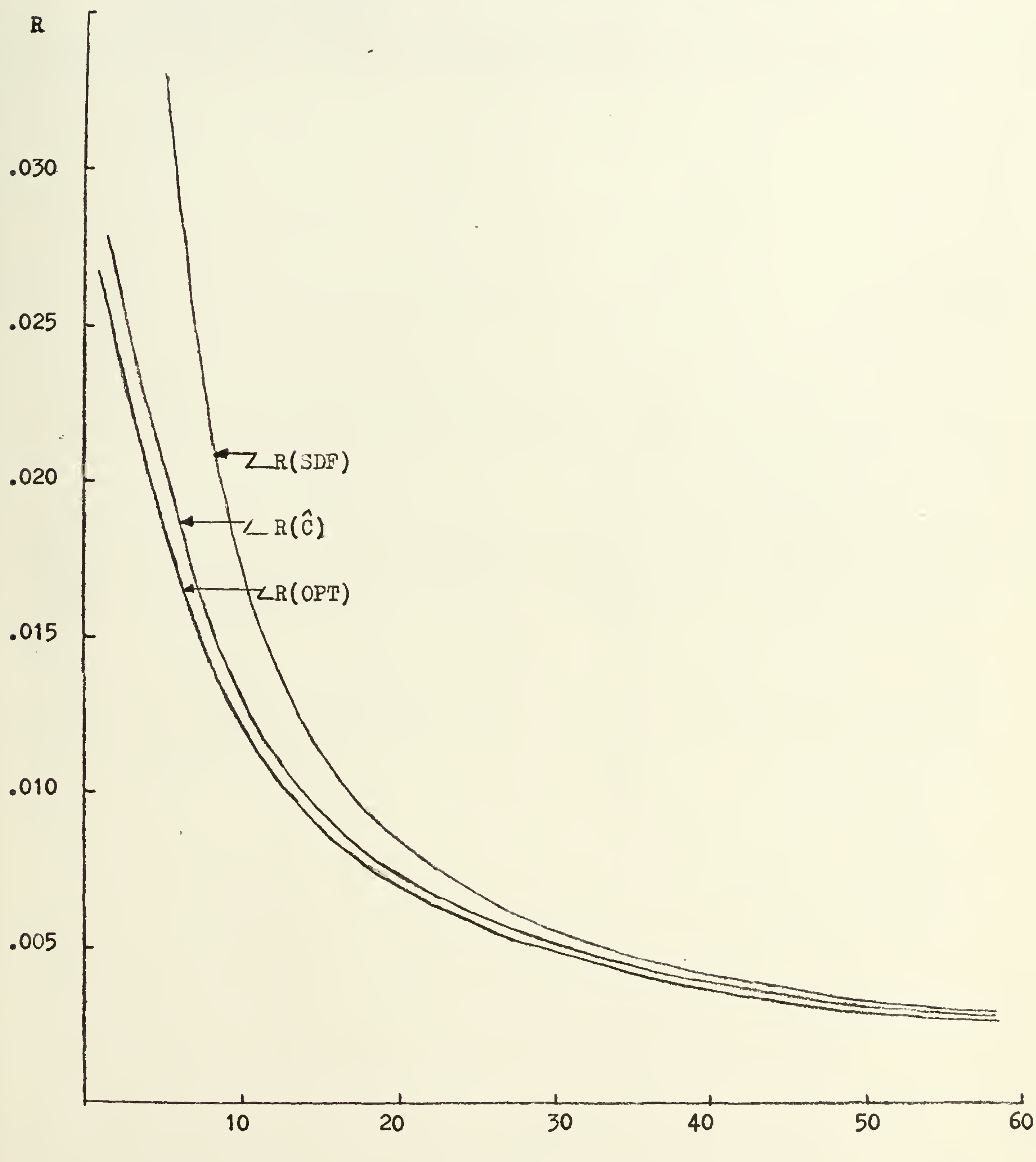
Cumputer output is shown on page 25 for n = 1 , . . . , 60 . Equations (4.1) , (4.2) and (4.3) are used .

N	(N+1)R($\hat{C}$)	(N+1)R(SCF)	(N+1)R(OPT)	LAST TWO TERMS OF (N+1)R(OPT)
1	0.05555555	0.33333330	0.05353007	-0.01649304
2	0.08333331	0.25000000	0.07628661	-0.00857831
3	0.09999996	0.22222220	0.09249049	-0.00582393
4	0.11111100	0.20833330	0.10353520	-0.00399830
5	0.11904750	0.19999990	0.11179110	-0.00294773
6	0.12499990	0.19444440	0.11814010	-0.00225484
7	0.12962950	0.19047610	0.12318110	-0.00178204
8	0.13333320	0.18750000	0.12727660	-0.00144335
9	0.13636360	0.18518510	0.13066930	-0.00119288
10	0.13888870	0.18333320	0.13352540	-0.00100234
11	0.14102550	0.18181810	0.13596240	-0.00085407
12	0.14285700	0.18055550	0.13806620	-0.00073642
13	0.14444440	0.17948710	0.13990060	-0.00064150
14	0.14583320	0.17857140	0.14151430	-0.00056382
15	0.14705870	0.17777770	0.14294460	-0.00049944
16	0.14814800	0.17708330	0.14422110	-0.00044549
17	0.14912270	0.17647050	0.14536740	-0.00039983
18	0.14999980	0.17592590	0.14640240	-0.00036084
19	0.15079360	0.17543850	0.14734160	-0.00032730
20	0.15151500	0.17499990	0.14819760	-0.00029822
21	0.15217380	0.17460310	0.14898100	-0.00027285
22	0.15277770	0.17424230	0.14970070	-0.00025059
23	0.15333320	0.17391300	0.15036420	-0.00023094
24	0.15384600	0.17361110	0.15097770	-0.00021352
25	0.15432090	0.17333320	0.15154670	-0.00019799
26	0.15476170	0.17307680	0.15207600	-0.00018410
27	0.15517230	0.17283940	0.15256950	-0.00017163
28	0.15555540	0.17261900	0.15303060	-0.00016038
29	0.15591390	0.17241370	0.15346260	-0.00015020
30	0.15624980	0.17222210	0.15386800	-0.00014095
31	0.15656550	0.17204290	0.15424930	-0.00013254
32	0.15686260	0.17187500	0.15460860	-0.00012486
33	0.15714280	0.17171710	0.15494760	-0.00011783
34	0.15740730	0.17156850	0.15526810	-0.00011137
35	0.15765750	0.17142850	0.15557160	-0.00010543
36	0.15789460	0.17129620	0.15585930	-0.00009996
37	0.15811950	0.17117110	0.15613250	-0.00009490
38	0.15833320	0.17105250	0.15639220	-0.00009021
39	0.15853650	0.17094010	0.15663940	-0.00008586
40	0.15873000	0.17083320	0.15687500	-0.00008182
41	0.15891460	0.17073160	0.15709970	-0.00007806
42	0.15909080	0.17063480	0.15731440	-0.00007455
43	0.15925920	0.17054250	0.15751970	-0.00007128
44	0.15942010	0.17045450	0.15771620	-0.00006821
45	0.15957430	0.17037030	0.15790440	-0.00006534
46	0.15972210	0.17028980	0.15808480	-0.00006265
47	0.15986380	0.17021270	0.15825800	-0.00006012
48	0.15999990	0.17013880	0.15842430	-0.00005773
49	0.16013060	0.17006800	0.15858420	-0.00005549
50	0.16025630	0.16999990	0.15873790	-0.00005338
51	0.16037730	0.16993460	0.15888600	-0.00005138
52	0.16049370	0.16987170	0.15902860	-0.00004950
53	0.16060590	0.16981130	0.15916600	-0.00004771
54	0.16071420	0.16975300	0.15929870	-0.00004603
55	0.16081860	0.16969690	0.15942670	-0.00004443
56	0.16091940	0.16964280	0.15955030	-0.00004291
57	0.16101680	0.16959050	0.15966980	-0.00004146
58	0.16111090	0.16954020	0.15978530	-0.00004009
59	0.16120200	0.16949140	0.15989710	-0.00003879
60	0.16129020	0.16944440	0.16000530	-0.00003755

After computer calculations for n = 1 , . . . , 60 (4.1) , (4.2) and (4.3) is shown in (4.4) and (4.5)



(4.4)



(4.5)

LIST OF REFERENCES

1. Read, R. R., The Asymptotic Inadmissibility of the Sample Distribution Function, Naval Postgraduate School, 1970.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Prof R. R. Read, Code 55Kd Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1
4.	LT(JG) Mete Ozer, Turkish Navy Sadri Maksudi Arsal sokak, No. 166-5 Kurtulus - Istanbul, TURKEY	1
5.	Genel Kurmay Baskanligi BAKANLIKLAR - ANKARA TURKEY	1
6.	Deniz Kuvvetleri Komutanligi BAKANLIKLAR - ANKARA TURKEY	1
7.	Deniz Harb Okulu Komutanligi HEYBELIADA - ISTANBUL TURKEY	1

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3 REPORT TITLE

A Problem in Estimating a Distribution by Minimum Risk

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)
Master's Thesis; September 1970

5 AUTHOR(S) (First name, middle initial, last name)

Mete Ozer

6 REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1970	30	1

8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

It is common to estimate a distribution by means of a step function. Such estimates can be made continuous by connecting the points of the steps with straight line segments. In this paper the best estimator of this class is found for data which is uniformly distributed using minimum risk, then this risk is compared with those of the sample distribution function and the Pyke estimator.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Estimating Distribution						
Uniform Distribution						

DD FORM 1 NOV 65 1473 (BACK)

S/N 0101-807-6821

A-31409

Thesis
0997
c.1

121664

Ozer

A problem in estimating a distribution by minimum risk.